

Variables Asociadas a la Satisfacción de Alumnos mediante Técnicas de Minería de Reglas de Asociación

Susana Beatriz Ruiz ; Myriam Beatriz Herrera

Facultad de Ciencias Exactas, Físicas y Naturales, Universidad Nacional de San Juan, Argentina.
sbruizr@yahoo.com.ar ; myriamhrrr@gmail.com

Resumen

Al igual que la mayoría de las organizaciones, las instituciones educativas deben evaluar la calidad de los servicios prestados. Aquí, el concepto de "calidad" se entiende como la forma en que esta organización satisface las expectativas de sus estudiantes para contribuir a la imagen que los estudiantes tienen de ella. Aunque la calidad percibida en el servicio recibido está fuertemente relacionada con las necesidades de los estudiantes, es posible extraer indicadores objetivos como su satisfacción como un factor esencial para abordar la calidad de la educación superior. Este artículo presenta los resultados de un estudio exploratorio sobre la satisfacción de 112 estudiantes de la Facultad de Ciencias Exactas, Naturales y Físicas de la Universidad Nacional de San Juan durante el año 2017. Mediante el uso de técnicas de Minería de Datos, basadas en lógica bivalente, descubrimos relaciones entre variables lingüísticas del conjunto de datos, bajo en análisis de reglas de la forma $P \rightarrow \text{Satisfacción}$. Consideramos que la información encontrada es valiosa para tomar decisiones institucionales con el fin de mejorar la calidad educativa.

Palabras clave: Satisfacción, Descubrimiento, Relaciones, Reglas Asociación.

Introducción

En la actualidad, la mayoría de las instituciones, empresas u organizaciones analizan la calidad de sus productos y/o servicios. De igual forma

las instituciones educativas se ven obligadas a valorar la calidad educativa. La calidad se debe entender como "un concepto abstracto tan amplio en definición y aplicación que cada organización puede entenderlo desde sus propios intereses" [1]. El concepto abarca el modo en que la empresa satisface todas las especificaciones de sus clientes y todo contacto con ellos, ya que contribuye a la imagen que se forman los clientes en sus mentes sobre la empresa [2]. La percepción de la calidad en el servicio debe estar respaldada por una estructura y gestión organizacional, de tal forma que se cree una filosofía de servicio que busque superar las expectativas de los clientes [3]. Cabe destacar que la calidad percibida está muy relacionada con la adecuación de las características del objeto a las necesidades del individuo. A pesar de esto, el contacto del individuo con el servicio que recibe y con los agentes que se lo ofrecen, pueden brindar datos objetivos [4].

Las instituciones educativas universitarias no deben estar ajenas a esta mirada. Los estudios sobre satisfacción estudiantil en las universidades como indicadores para evaluar la calidad educativa son particularmente importantes, porque la satisfacción estudiantil mejora el rendimiento académico, reduce la deserción, el cambio de carrera de los inscritos y es un requisito para el éxito en el aprendizaje. A su vez, una comprobada calidad educativa fortalece la imagen y el prestigio de la institución [5]. Valorar la satisfacción estudiantil y determinar cuáles son las variables asociadas en este aspecto, contribuye a la toma de decisiones acertadas en la gestión orientada a la calidad universitaria.

La disponibilidad de grandes volúmenes de información y el uso generalizado de herramientas informáticas, hoy en día, ha transformado el análisis de datos orientándolo hacia determinadas técnicas especializadas englobadas bajo el nombre de Minería de Datos o Data Mining (DM). Las técnicas de DM persiguen el descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en grandes bases de datos. Tiene como el objetivo de detectar patrones de comportamiento consistentes o relaciones entre las diferentes variables para aplicarlos a nuevos conjuntos de datos [6]. Para realizar este proceso no trivial de identificación válida, novedosa, potencialmente útil y entendible de patrones comprensibles que se encuentran ocultos en los datos se utilizan técnicas estadísticas que permiten reunir y analizar los datos, como también técnicas provenientes de la inteligencia artificial así como redes neuronales, algoritmos genéticos, técnicas de lógica fuzzy y otros tipos de heurísticas aplicados a los procesos de extracción de patrones ocultos o información útil de los conjuntos de datos. Tanto las técnicas estadísticas como las técnicas de inteligencia artificial son bastante poderosas; en algunos casos, estas son tan solo dos enfoques o alternativas diferentes para la solución de un mismo problema, en otras ocasiones son técnicas complementarias porque resuelven problemas de naturaleza diferentes [7].

Dentro de ella, la MD, se encuentran estrategias de descubrimiento de reglas de asociación, que tienen como base la lógica clásica, y se proponen encontrar conjuntos de elementos que co-ocurren juntos frecuentemente en una base de datos [8] [9] [10]. La aplicación de esta tecnología, también denominada Minería de Reglas de Asociación (MRA), puede generar una gran cantidad de reglas, por lo que surge la necesidad de seleccionar aquellas reglas de asociación relevantes desde la perspectiva específica de estudio.

Las reglas de asociación son una de las técnicas más conocidas dentro de la minería de datos. Este tipo de herramientas permiten encontrar relaciones frecuentes de aparición conjunta de

objetos o ítems existentes en una base de datos. La gran ventaja de estas técnicas es que permite encontrar, dentro de un rango de medidas de significancia establecidas previamente por el analista, múltiples relaciones y mostrarlas en forma de reglas de conocimiento del tipo “SI ... ENTONCES ...”. Este conocimiento puede ser analizado por los expertos del dominio con el objetivo de encontrar relaciones importantes, no triviales y desconocidas previamente [11].

Se puede afirmar que, en general, un indicador directo de la calidad educativa es la satisfacción del alumno por la carrera [12][13][14][15]. La satisfacción del estudiantado universitario constituye un factor imprescindible en el abordaje del tema de la calidad de la educación superior. Vista la importancia del tema, en este trabajo se muestran resultados de un estudio exploratorio sobre la satisfacción de 112 estudiantes que cursan carreras en la Facultad de Ciencias Exactas, Naturales y Físicas de la Universidad Nacional de San Juan (Argentina), en el año 2017, con el propósito de determinar las variables asociadas que mejor caracterizan la satisfacción del alumnado por carreras que cursan, aplicando técnicas de DM, basados en lógica bivalente. Los análisis se centran en el descubrimiento de relaciones de la forma $P \rightarrow \text{Satisfacción}$, entre estados lingüísticos a partir de un conjunto de datos proveniente de una encuesta.

Metodología

Se quiere caracterizar la satisfacción del alumnado por carreras de la FCEFyN de la UNSJ (Argentina) mediante la determinación de variables asociadas, a partir de datos provenientes de una encuesta en la que participaron 112 alumnos en el año 2017. La encuesta trata sobre Factores de Riesgo y Calidad de Vida y fue elaborada con la herramienta web de encuestas online EncuestaFácil.com, y cuenta con varias secciones. Las variables consideradas se pueden agrupar en: variables que caracterizan la Facultad, Universidad y la carrera/s que cursa el estudiante; variables que representan

características personales del estudiante y de su familia (edad, sexo, su relación con pares, etc.). Variables asociadas al rendimiento (rendimiento, promedio, etc.); y variables que representan el esfuerzo y motivación del estudiante (Ej. hs. de estudio, asistencia a la universidad, la percepción sobre la carrera en cuanto a que mejorará las expectativas en el futuro, el grado de satisfacción por la carrera, etc.). Se puede consultar la encuesta en: <https://www.encuestafacil.com/RespWeb/Qn.aspx?EID=2197195>. Las preguntas consideradas no son mutuamente excluyentes entre sí. Por lo que cada pregunta es una variable en sí misma; las respuestas alternativas a las preguntas de cada sección sí son mutuamente excluyentes y cada una de estas respuestas es una modalidad (ítems, categoría o etiqueta lingüística) de las variables cualitativas a la que pertenece. La información resultante es pre procesada convenientemente, utilizando el software Excel, para la aplicación de los distintos procedimientos que posteriormente se especifican. En este estudio, a partir del cuestionario, se consideraron un total de 16 variables lingüísticas y 56 estados lingüísticos, ítems o categorías, para realizar los estudios.

Dada la importancia de la temática a abordar y la naturaleza de las variables involucradas en la encuesta, con el propósito de descubrir relaciones relevantes entre variables lingüísticas de la forma lógica $P \rightarrow Satisfacción$. En este trabajo se tiene en cuenta resultados de un análisis preliminar, con los datos de encuesta, en la que se observa la existencia de variables y estados lingüísticos cuyas implicaciones resultan verdaderas en un porcentaje mayor al 50% de casos. Se decide por consiguiente aplicar un procedimiento clásico de DM, el algoritmo de búsqueda Apriori, para descubrir relaciones entre variables y definir reglas de asociación. La tarea se realiza con la ayuda del software libre R, versión 3.5.2.

Entre los procedimientos basados en la lógica clásica, dentro de la MD, el problema de descubrir asociaciones a partir de los datos

consiste en identificar grupos de variables que se correlacionan fuertemente entre sí. Se cuenta con un conjunto de ítems y un gran conjunto de transacciones que son subconjuntos de esos ítems. La Minería de Reglas de Asociación (MRA) procura encontrar relaciones entre ítems, llamadas Reglas de Asociación, a partir de la presencia frecuente de varios ítems dentro de las transacciones. Formalmente, se consideran los siguientes conceptos para abordar la problemática de búsqueda de reglas de asociación: $D = \{d_1, d_2, \dots, d_m\}$ es un conjunto de ítems; $T = \{t_1, t_2, \dots, t_n\}$ es un conjunto de transacciones, donde cada transacción t_i es un conjunto de ítems tal que $t_i \subseteq D$, $1 \leq i \leq n$. La implicación $X \rightarrow Y$ es una *Regla de Asociación* donde $X \subset D$, $Y \subset D$, $X \cap Y = \emptyset$ y $X \cup Y \subseteq t_i$. Es decir, los conjuntos X e Y son mutuamente excluyentes, t_i es el conjunto de ítems formado por aquellos que corresponden al antecedente o al consecuente de la Regla de Asociación. El conjunto $X \cup Y$ debe estar contenido o ser igual a alguna de las transacciones pertenecientes a T . Un procedimiento básico para hallar itemset frecuentes para obtener reglas de asociación booleanas es el algoritmo Apriori. El algoritmo emplea una aproximación iterativa conocida como búsqueda inteligente de nivel, donde los k-itemset se utilizan para explorar el k+1 nivel. Para mejorar la eficacia emplea la propiedad a priori que indica que todos los subset de un itemset frecuente deben ser frecuentes. Lo que indica que si un conjunto no supera una prueba los supra-conjuntos derivados de este tampoco lo superarán.

En este trabajo el conjunto D está conformado por el conjunto de “respuestas sí” (ítems) a responder en la encuesta; mientras que cada transacción se identifica como “conjunto de respuestas sí” (conjuntos de ítems) de un cierto alumno. Hay tantos ítems como “respuestas sí” del cuestionario y tantas transacciones como alumnos encuestados.

Una vez que los itemset frecuentes de las transacciones en la base de datos son hallados

se pasa a seleccionar las reglas de asociación fuertes. Las medidas más populares en MRA son soporte, confianza y *lift* [16][17]. En el marco soporte-confianza, originariamente [18][19][20], la búsqueda de reglas de asociación adopta los factores soporte y confianza para evaluar las reglas descubiertas. El soporte de un ítem es la frecuencia con la cual este ítem se encuentra en las transacciones dividido entre el número de transacciones. Esto es, si X es un ítem entonces:

$$\text{Soporte}(X) = \frac{\text{n}^\circ \text{ de transacciones que contienen al ítem } X}{\text{n}^\circ \text{ total de transacciones de la base de datos}} \quad (1)$$

Para obtener el soporte de una regla de asociación $X \rightarrow Y$, se obtiene con la siguiente ecuación:

$$\text{Soporte}(X \rightarrow Y) = \frac{\text{n}^\circ \text{ de transacciones que contienen los ítems } X \text{ y los ítems } Y}{\text{n}^\circ \text{ total de transacciones de la base de datos}}$$

La medida de confianza de una regla de asociación $X \rightarrow Y$ es la división entre el soporte de la regla de asociación y el soporte del antecedente de la regla de asociación, esto está representado por la siguiente ecuación:

$$\text{Conf}(X \rightarrow Y) = \frac{\text{Soporte}(X \rightarrow Y)}{\text{Soporte}(X)} \quad (2)$$

La confianza puede interpretarse como un estimador de $P(Y/X)$, la probabilidad de encontrar la parte derecha de una regla condicionada a que se encuentre también la parte izquierda [21].

La confianza está dirigida y computa valores diferentes para las reglas $X \rightarrow Y$ e $Y \rightarrow X$.

Según Romero [22] la confianza no es capaz de detectar la independencia estadística. Este mismo autor expresa que normalmente se ha creído que mientras mayor es el soporte, mejor es el conjunto de elementos, pero esto puede no ser siempre cierto ya que un conjunto de elementos con alto soporte puede ser fuente de engaños, debido a que aparecen en la mayoría de las transacciones. Las restricciones que pueden presentar los factores de soporte y confianza, unidos a la necesidad de rescatar las

reglas interesantes de un conjunto posiblemente numeroso de reglas generadas, dificulta la actividad de los expertos en el campo en el que se aplica MD. A raíz de ello, distintos autores han desarrollado otras medidas para evaluar la importancia de las reglas generadas. Entre estas medidas se rescata la denominada *lift* o medida de independencia [19][20][22]. El factor *lift* representa una prueba para medir la dependencia estadística y se define como:

$$\text{lift}(X \rightarrow Y) = \text{lift}(Y \rightarrow X) = \frac{P(X \cup Y)}{P(X) \cdot P(Y)} \quad (3)$$

Este factor establece una relación entre la ocurrencia simultánea de X e Y , cuando los conjuntos de ítems que conforman el antecedente y el consecuente de la regla sean estadísticamente independientes. Como el *lift* es simétrico, este valor sólo mide el grado de dependencia y no la implicación en ambas direcciones.

Según Hahsler [10] las reglas recuperadas por medio de soporte y confianza, deberían ser filtradas usando sus valores de *lift*, pues valores de *lift* mayores a 1 indican asociación entre ítems, mientras que valores menores a 1 pueden indicar su independencia. Reglas de asociación con valores de *lift* menores a 1 no deberían ser tenidas en cuenta para la toma de decisiones.

Para aplicar el procedimiento, utilizando el software R, se requiere de la implementación previa de los paquetes “arules” y “arulesViz” [23].

El programa que se confecciona sigue en términos generales los siguientes pasos:

- 1) Lectura de datos de encuesta e instalación de los paquetes “arules” y “arulesViz”.
- 2) Construcción de transacciones.
- 3) Aplicación de la función “apriori” para la definición de los itemsets frecuentes.
- 4) Construcción de gráficos para explorar itemset, con la ayuda del comando “plot” y los métodos “graph” y “scatterplot”.
- 5) Filtrado de itemsets frecuentes de la forma $P \rightarrow \text{Satisfacción}$, mediante la aplicación de la función “apriori” considerando un valor mínimo de


soporte (70% de los datos) y confianza mínima de 0.7.

- 6) Selección de reglas de asociación fuertes mediante evaluación de soporte, confianza y lift.

Resultados

A continuación, se presentan los resultados obtenidos, según la metodología de trabajo detallada en la sección anterior. Los resultados se analizan desde las perspectivas específicas de estudio y teniendo en cuenta el objetivo planteado en el presente trabajo.

Se quieren encontrar relaciones entre ítems de la forma $X \rightarrow [Satisfacción_carrera_Si]$, dentro de un conjunto de transacciones definidas por las respuestas de alumnos. Luego de la lectura de datos de encuesta e instalación de los paquetes “arules” y “arulesViz”, en R, se definen las transacciones considerando toda la información proveniente a la base de datos, para posteriormente aplicar la función “apriori”. En la Figura 1 se pueden observar 21 ítems entre los considerados en este estudio, en la cual los 19 primeros son parte de la transacción 1.



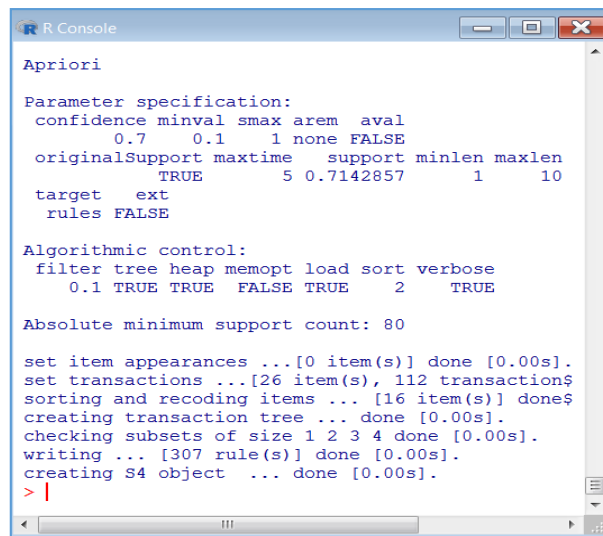
```
> datos=read.table("C:/Prueba.csv",h=TRUE)
> library(arulesViz)
> library(arules)
> datos

      Id_alumno.Items
1  1,Vivienda_adeCuada_estudio
2  1,dispositivos_electricos=Si
3  1,Acceso_internet_vivienda_Si
4  1,Dinero_Estudio_suficiente
5  1,AsistenciaXsemana_Mayoria_dias
6  1,Promedio_Bueno
7  1,Efecto_Aplazos_si
8  1,Nivel_exigencia_carrera_alto
9  1,Tiempo_dedicado_teoria_Mas_de4hs
10 1,Tiempo_practica_mas4hs
11 1,Tiempo_estudio_solo_mas4hs
12 1,Accesibilidad_Material_bueno
13 1,Satisfaccion_carrera_Si
14 1,Relacion_compañeros_buena
15 1,Relacion_docentes_buena
16 1,Estrés_estudio_si
17 1,carrera_mejora_futuro_Si
18 1,Rendimiento_según_expectativas_Bueno
19 1,Vivienda_adeCuada_estudio
20 2,AsistenciaXsemana_Mayoria_dias
21 2,Promedio_Bueno
```

Figura 1

Como se observa en la Figura 2, se hallaron 307 reglas de asociación generales para la base de datos considerada, con valor mínimo de soporte de 0.71. Las Figuras 3 y 5 ilustran el ranking de

ítems más frecuentes, dentro de los 20 primeros. Mientras que el grafo de la Figura 4 proporciona una caracterización gráfica de los valores de confianza, soporte y lift hallados, para el conjunto de las primeras 40 reglas generales descubiertas con mayores valores de confianza, entre las que incluye el ítem “satisfacción del alumno= si”.



```
Apriori

Parameter specification:
confidence minval smax arem aval
0.7 0.1 1 none FALSE
originalSupport maxtime support minlen maxlen
TRUE 5 0.7142857 1 10
target ext
rules FALSE

Algorithmic control:
filter tree heap memopt load sort verbose
0.1 TRUE TRUE FALSE TRUE 2 TRUE

Absolute minimum support count: 80

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[26 item(s), 112 transaction$
sorting and recoding items ... [16 item(s)] done$
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [307 rule(s)] done [0.00s].
creating S4 object ... done [0.00s].
> |
```

Figura 2

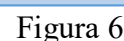
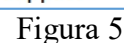
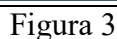
Las Figuras 6 y 7 muestran resultados luego del filtrado de itemsets frecuentes, considerando como valor mínimo de soporte 0.7 y valor mínimo de confianza 0.7. Se obtienen un total de 10 reglas de asociación de la forma $X \rightarrow [Satisfacción_carrera_Si]$. Entre las 10 reglas obtenidas, se seleccionan aquellas más fuertes (con valores de lift mayores a 1).

- {2}[Estrés _ estudio _ si] \rightarrow [Satisfacción _ carrera _ Si]
- {3}[Carrera _ mejor _ futuro _ si] \rightarrow [Satisfacción _ carrera _ Si]
- {4}[Relación _ Compañeros _ Buena] \rightarrow [Satisfacción _ carrera _ Si]
- {5}[Promedio _ Bueno] \rightarrow [Satisfacción _ carrera _ Si]
- {7}[Carrera _ mejora _ futuro _ Si, Promedio _ Bueno] \rightarrow [Satisfacción _ carrera _ Si]
- {8}[Promedio_Bueno, Relación _ Compañeros _ buena] \rightarrow [Satisfacción _ carrera _ Si]

Para la regla “{2}” obtenida, el valor de soporte expresa que en el 71,4% de las transacciones estuvieron involucrados alumnos que expresan

El valor de lift mayor a uno, indica asociación (dependencia estadística) entre los ítems “Estrés_estudio_si” y “Satisfacción_carrera_Si”. En forma similar se pueden interpretar las reglas “{3}”, “{4}”, “{5}”, “{7}” y “{8}”.

Asumiendo que la muestra es representativa de la población de alumnos de la institución educativa, se puede concluir que: desde el estudio realizado de reglas de asociación, utilizando el algoritmo Apriori implementado en R, se han encontrado evidencias para afirmar que los ítems “estrés por el estudio”, “percepción del alumno que la carrera mejorará su futuro”, “buena relación con compañeros” y “promedio bueno en la carrera” están asociados al ítem “satisfacción del alumnado por carreras”.



Conclusiones

Este trabajo muestra un informe que resulta de un análisis exploratorio de datos proveniente de una encuesta, que resulta de aplicar técnicas de Minería de Datos. Con el propósito de caracterizar al alumnado, en cuanto a la satisfacción por las carreras que cursan, se aplican técnicas de descubrimiento de relaciones entre variables involucradas en reglas de asociación de la forma $P \rightarrow \text{Satisfacción}$.

Desde la aplicación del algoritmo Apriori implementado en R, se pudo encontrar reglas de asociación fuertes que permiten asociar estadísticamente las variables: “Estrés por el estudio”, “Mejoras en el futuro con la carrera”, “Buena relación con compañeros” y “Buen promedio en la carrera” con la variable “Satisfacción del alumnado por carreras”. Estas muestran que la valoración de la satisfacción del alumnado y de la calidad educativa en las carreras impartidas, desde la perspectiva del alumnado, están vinculadas con aspectos personales, actitudinales, institucionales, económicos, sociales y culturales.

Consideramos que la metodología aplicada en este trabajo resulta ser valiosa para extraer información significativa y que puede ayudar a la hora de la toma de decisiones institucional, para la mejora de la calidad educativa.

Queda como trabajo a futuro complementar este trabajo, con otras técnicas de Minería de Datos, como son las técnicas estadísticas de clasificación (análisis de conglomerados y análisis de discriminantes), a fin de definir patrones o modelos que permiten discriminar al alumnado según su grado de satisfacción por las carreras. También indagar sobre otras posibles asociaciones entre variables consideradas en la encuesta indicadoras de la calidad educativa, como por ejemplo es el rendimiento académico del alumnado según sus expectativas, empleando las metodologías citadas; y/o técnicas basadas en Lógicas Multivalentes (entre ellas la Lógica Fuzzy) en el caso de que

la naturaleza de las variables de estudio lo permita.

Bibliografía

- [1] Traba, L., Barletta, M. y Velázquez, J. (2010). *Teoría y práctica de las organizaciones: herramienta para la gestión de la calidad*. Santa Fe: Universidad Nacional del Litoral.
- [2] Fernández Ziegler, R. (2003). *Planificación y Control de Gestión*. Buenos Aires: Editorial de la Universidad de Quilmes.
- [3] Botero, M.M. y Peña, P. (2006). *Calidad en el servicio: el cliente incógnito*. Suma Psicológica, 13(2), 217-228.
- [4] C. Salvador, C.; Pozo C.; Alonso, E. (2008)- *Percepción del cliente de los predictores de calidad en el sector servicios*- Boletín de Psicología, No. 94, Noviembre 2008, 69-84. Recuperado de: <https://www.uv.es/seoane/boletin/previos/N94-5.pdf>
- [5] Surdez, E. G., Sandoval, M del C. y Lamoyi, C. L. (2018). *Satisfacción estudiantil en la valoración de la calidad educativa universitaria*. Educación y Educadores, 21(1), 9-26. DOI: 10.5294/edu.2018.21.1.1 Satisfacción estudiantil en la valoración de la calidad educativa universitaria]
- [6] Marín Llanes Luis A., Carro Cartaya Juan Carlos "La Minería de Datos como Herramienta en el Proceso de Inteligencia Competitiva" Consultoría Biomundi, Dirección de Inteligencia Corporativa, Instituto de Información Científica y Tecnológica (IDICT), CUBA, Taller Nacional sobre Inteligencia Empresarial IntEmpres'2000 Recuperado de: <http://www.redciencia.cu/empres/index.htm>
- [7] ROJAS DIAZ, J.; CHAVARRO PORRAS, J.; MORENO LAVERDE, R(2009). *Técnicas de lógica difusa aplicadas a la minería de datos*. Scientia et technica, [S.l.], v. 3, n. 40, ISSN 2344-7214. Recuperado de: <http://revistas.utp.edu.co/index.php/revistaciencia/article/view/3095/1695>>. doi:<http://dx.doi.org/10.22517/23447214.3095>.

- [8] AGRAWAL, R., IMIELINSKI, T., SWAMI, A.(1993) *Mining association rules between sets of items in large databases*. In: Proceedings of the ACM SIGMOD International Conference on Management of Data. Washington D.C., p. 207-216.
- [9] HIPPI, J.; GÜNTZER, U.; NAKHAEIZADEH, G.(2000). *Algorithms for association rule mining—a general survey and comparison*. ACM sigkdd explorations newsletter, vol. 2, no 1, p. 58-64.
- [10] HAHLER, Michael; HORNIK, Kurt; REUTTERER, Thomas (2006). *Implications of probabilistic data modeling for mining association rules*. En From Data and Information Analysis to Knowledge Engineering. Springer Berlin Heidelberg. p. 598-605.
- [11] MARCHÁN, E.; SALCEDO, J.; AZA, T.; FIGUERA, L.; MARTÍNEZ DE PISÓN, F.; GUILLÉN, P. (2011) Reglas de asociación para determinar factores de riesgo epidemiológico de transmisión de la enfermedad de Chagas. *Revista Ciencia e Ingeniería*, pp.55-60. Recuperado de: https://www.researchgate.net/publication/273134305_Reglas_de_asociacion_para_determinar_factores_de_riesgo_epidemiologico_de_transmision_de_la_enfermedad_de_Chagas [accessed Jun 23 2019].
- [12] ALVES, H. y RAPOSO, M. (2004). *La medición de la satisfacción en la enseñanza universitaria: El ejemplo de la Universidad de Da Beira Interior*. International Review on Public and Nonprofit Marketing, Portugal, 1(1): 73-88.
- [13] SALINAS A. y MARTÍNEZ, P. (2007). Principales factores de satisfacción entre los estudiantes universitarios. *Revista Internacional de Ciencias Sociales y Humanidades*. México, XVII (001):163-192.
- [14] SALINAS, A. (2007). Satisfacción del estudiante y calidad Universitaria: Un análisis explicatorio en la Unidad Académica Multidisciplinaria Agronomía y Ciencias de la Universidad Autónoma de Tamaulipas. Tesis Doctoral, Universidad de Sevilla.
- [15] FERNÁNDEZ, J.; FERNÁNDEZ, S; ÁLVAREZ, A.; MARTÍNEZ, P. (2007). Éxito Académico y Satisfacción de los Estudiantes con la Enseñanza Universitaria. *Revista Electrónica de Investigación y evaluación Educativa*. España, 13(2):203-214.]
- [16] BROWN, M. (2014) *Data mining for dummies*. John Wiley & Sons.
- [17] LUCAS, Joel Pinho .(2010) *Métodos de clasificación basados en asociación aplicados a sistemas de recomendación*. Tesis Doctoral. Universidad de Salamanca.
- [18] AGRAWAL, R., IMIELINSKI, T., SWAMI, A. Mining association rules between sets of items in large databases. In: Proceedings of the ACM SIGMOD International Conference on Management of Data. 1993. Washington D.C., p. 207-216.
- [19] BRIN, S., MOTWANI, R., ULLMAN, J. D., TSUR, S. Dynamic itemset counting and implication rules for market basket data. Proceedings ACM SIGMOD International Conference on Management of Data. 1997. Tucson, Arizona, USA. p. 255-264.
- [20] SILVERSTEIN, Craig; BRIN, Sergey; MOTWANI, Rajeev. Beyond market baskets: Generalizing association rules to dependence rules. Data mining and knowledge discovery, 1998, vol. 2, no 1, p. 39-68.
- [21] Hipp, J.; Ulrich Güntzer, U.; Nakhaeizadeh, G..(2000) Algorithms for association rule mining - A general survey and comparison. SIGKDD Explorations, 2(2):1-58.
- [22] ROMERO, C.(2003) Aplicación de técnicas de adquisición de conocimiento para la mejora de cursos hipermedia adaptativos basados en Web. Tesis Doctoral. Universidad de Granada. E.T.S. Ingeniería Informática. .
- [23] AMAT, R. (2018). Reglas de asociación y algoritmo Apriori con R. Recuperado de: https://rpubs.com/Joaquin_AR/397172